

QUANTITATIVE STILMERKMALE

Es dürfte bekannt sein, daß quantitative Größen wie Wort- und Satzlänge, Worthäufigkeitsverteilungen und andere Parameter wie Häufigkeit der einzelnen Wortklassen und grammatischen Formen als geeignete Stilmerkmale zur Charakterisierung eines Autors und verschiedener Texte verwendet werden können. Seit der Arbeit von G.U. Yule, *The Statistical Study of Literary Vocabulary*, Cambridge 1944, ist die Stilanalyse auf mathematisch-statistischer Grundlage für die Beschreibung der Formalstruktur eines Textes ein wichtiges methodisches Instrument geworden.¹

Es dürfte ebenso bekannt sein – muß aber wieder betont werden – daß eine erschöpfende Textbeschreibung quantitative und qualitative Methoden verbinden muß. Man kann allerdings feststellen, daß sowohl in linguistischen als auch literaturwissenschaftlichen Untersuchungen auch einfache glossodynamische Parameter, die sich für strukturelle Textidentifikationen und Textsortenvergleiche sehr gut eignen, noch verhältnismäßig wenig verwendet worden sind.

Im folgenden werden wir anhand von sechs quantitativen Stilmerkmalen zwölf Texte analysieren. Es sind sechs Flugblätter (=Fb), die in den Jahren 1972-73 in der Universität Hamburg verbreitet worden sind und sechs Texte der Wahlwerbung (=Ww) (5 SPD, 1 CDU, Texte in Zeitungen und Zeitschriften) aus dem Jahr 1973. Die Texte wurden durch das random-sample Verfahren aus einer Menge von 30 Texten jeder Gruppe ausgewählt.²

Das Ziel dieser Untersuchung ist es festzustellen, ob und wie sich die Textgruppen, Flugblätter und Wahlwerbung, unterscheiden, die nach ihren Funktionen beide als Werbetexte gekennzeichnet werden können. Gibt es Stilindikatoren, wodurch die beiden Gruppen sich unterscheiden? Für weitere Vergleiche und Gruppierungen bieten sich vor allem die Stilindikatoren Wortlänge und Satzlänge an, die von Fucks, Arens und anderen für verschiedene Prosaveröffentlichungen der deutschen Sprache verwendet worden sind.

Die quantitativen Stilmerkmale in der vorliegenden Analyse sind:

- 1) Die Wortlänge, gemessen durch die mittlere Zahl der Silben pro Wort (\bar{i}). Als Wort gilt eine selbständige zusammenhängende Folge von Graphemen.
- 2) Die Satzlänge, gemessen durch die mittlere Zahl der Wörter pro Satz (\bar{j}). Als Satz gilt eine Folge von Wörtern, die entweder texteinleitend oder nach den Graphemen (.), (!) und (?) anfängt und mit diesen Graphemen endet.
- 3) Die mittlere Wortlänge und mittlere Satzlänge zugleich ($\xi = \bar{i} \cdot \bar{j}$).
- 4) Der Adjektiv-Wort-Quotient, ermittelt anhand der Zahl der Adjektive je Wort (A/W).
- 5) Der Adjektiv-Verb-Quotient, ermittelt anhand der Zahl Adjektive je Verb (A/V). Aus Vergleichsgründen wird dieser Quotient mit 100 multipliziert.
- 6) Der Adjektiv-Satz-Quotient, ermittelt anhand der Zahl der Adjektive je Satz (A/S).

Die Zahl der Wörter in den Texten liegt zwischen 305 und 1526. Tabelle 1 zeigt die ersten drei Parameter für die Flugblätter, Tabelle 2 für die Wahlwerbung.

Text	\bar{i}	\bar{j}	$\bar{i} \cdot \bar{j}$
A. ASTA-INFO. Behörde plant Abschaffung der Drittelparität	2,434	29,166	70,909
B. F-A aktiv gegen demokratische Rechte	2,064	21,492	44,359
C. Warum noch eine Demonstration	2,217	28,740	63,525
D. ASTA bricht eigene Beschlüsse	2,107	19,409	40,894
E. Sofortige Aufnahme des Streiks	1,982	23,300	46,180
F. Entwicklungsplan der Universität	2,256	25,533	57,598
	13,060	147,640	323,465

M = 2,177 24,606 53,711

Tabelle 1 : Mittlere Wortlänge \bar{i} , mittlere Satzlänge \bar{j} ,
 ζ -Werte $\bar{i} \cdot \bar{j}$ für Flugblätter

Text	\bar{i}	\bar{j}	$\bar{i} \cdot \bar{j}$
1. Betriebs-Information	2,053	12,511	25,715
2. Wir führen das Gespräch mit dem Wähler	1,960	12,607	24,714
3. Auf keinen Fall Arbeitslosigkeit (Helmut Schmidt)	1,949	24,500	47,748
4. SPD-Sozialdemokraten	1,936	18,438	35,686
5. Das muß jedermann wissen über die Rentenversicherung (Walter Arendt)	2,247	15,487	34,793
6. Liebe Mitbürgerinnen und Mitbürger (Willy Brandt)	1,947	11,719	22,812
	12,092	95,262	191,500

M = 2,015 15,877 31,917

Tabelle 2 : Mittlere Wortlänge \bar{i} , mittlere Satzlänge \bar{j} ,
 ζ -Werte $\bar{i} \cdot \bar{j}$ für Wahlwerbung

Ehe wir zur Analyse dieser Tabellen gehen, geben wir als Vergleich Mittelwerte an, die durch einige andere Untersuchungen gewonnen worden sind.

Untersuchte Texte		\bar{i}	\bar{j}
Fucks	1. Dichter	1,689	18,211
	2. Schriftsteller	1,984	24,385
	3. Sämtliche Texte (Dichter + Schrift- steller)	1,815	21,55
Arens	Deutsche Erzähl- prosa	1,721	25,40

Tabelle 3 : Mittlere Wortlänge \bar{i} und mittlere Satzlänge \bar{j}
nach den Untersuchungen von Fucks und Arens.

Unter Dichtern faßt Fucks schönliteräre Schriftsteller wie Bergengruen, Carossa, Chamisso, Fontane, Fallada, Goethe, u.a. zusammen, unter Schriftstellern Wissenschaftler verschiedener Fächer, sowie auch Journalisten und Politiker. Natürlich würden sich Verschiebungen ergeben, wenn man andere Samples von schönliterären Schriftstellern hätte, wie auch ein Vergleich der Daten von Fucks und Arens erweist.³ Wie sehr sich die Daten von Corpus zu Corpus verschieben können, zeigt die mittlere Satzlänge 16 W/S, die Eggers anhand einer Auswahl von 50 populärwissenschaftlichen Autoren der Rowohlt's deutschen Enzyklopädie ermittelt hat.⁴ Fucks Behauptung betreffs der Mittelwerte anhand sämtlicher 50 von ihm untersuchten Texte, daß "die regellose Zuziehung zahlreicher weiterer Texte hieran vermutlich nicht allzuviel ändern wird"⁵, kann daher nur bedingte Geltung haben. Bei einer methodisch stringenteren Auswahl der Texte kann man erwarten zu Ergebnissen zu kommen, wo dies zutreffen könnte; weitere Arbeiten mit repräsentativen Corpora sind daher notwendig. Immerhin erhalten wir aber schon anhand der obigen Daten gewisse Bezugspunkte über die Wort- und Satzlänge in der deutschen Prosa.

Die beiden Textgruppen Flugblätter und Wahlwerbung zeigen anhand der Parameter in Tabelle 1 und 2 folgende Charakteristika.

Die mittlere Wortlänge \bar{i} weist nur einen geringen Unterschied zwischen den Fb und Ww auf: bei $\bar{i}_{Fb} = 2,177$ und $\bar{i}_{Ww} = 2,015$ ist die durchschnittliche Wortlänge in den Flugblättern rund 8 % höher als in der

Wahlwerbung. Betrachten wir die mittleren Satz­längen $\bar{j}_{Fb} = 24,606$ und $\bar{j}_{Ww} = 15,877$, so sehen wir, daß die Sätze in den Flugblättern um rund 56 % länger als in der Wahlwerbung sind. Verglichen mit den Zahlen $\bar{i} = 1,815$ und $\bar{j} = 21,55$, Tabelle 3, Dichter und Schriftsteller, ist die Wort­länge der Flugblätter um rund 18 % und die der Wahlwerbung um rund 9 % höher als der Mittelwert für die deutsche Prosa ermittelt durch Fucks.

Der Parameter $\xi = \bar{i} \cdot \bar{j}$ ergibt für die Flugblätter 53,711 und für die Wahlwerbung 31,917; die Zahl ist bei den Flugblättern rund 67 % höher als bei der Wahlwerbung. Bei derartigen Befunden ist augenscheinlich, daß man folgende Feststellung von Fucks zu berücksichtigen hat: "Extrem große Werte der mittleren Wortlänge \bar{i} deuten im allgemeinen auf Wort­schwulst, extrem große Werte der mittleren Satz­länge \bar{j} auf Satzschwulst. Tritt beides zusammen ($\xi = \bar{i} \cdot \bar{j}$ groß), so wird man einen Text vor sich haben, der im allgemeinen stilistisch unbefriedigend sein wird".⁶ Aus­sagen wie "stilistisch unbefriedigend" müssen allerdings präzisiert werden. Diese Frage, auch wenn es sich um die Formalstruktur handelt, kann nur anhand verschiedener Normen gelöst werden, die uns bis jetzt nicht vorliegen.

Tabelle 4 gibt die Werte $\bar{i} \cdot \bar{j}$ in steigender Ordnung für alle Texte.

Text	$\bar{i} \cdot \bar{j}$	Text	$\bar{i} \cdot \bar{j}$
6	22,812	B	44,359
2	24,714	E	46,180
1	25,715	3	47,748
5	34,793	F	57,598
4	35,686	C	63,525
D	40,894	A	70,909

Tabelle 4 : ξ -Werte in steigender Ordnung

Die ξ -Werte für die Flugblätter liegen zwischen rund 40 und 70, für die Wahlwerbung zwischen rund 22 und 35. Beide Textgruppen unterscheiden sich dadurch deutlich als zwei Blöcke voneinander, vgl. Tabelle 4, Text 6 - 4 und Text D - A. Die extremen Werte (Text 6 und Text A) sind hier vor allem von Interesse. Folgende Textproben mögen die Struktur beleuchten:

Text 6, mit einem extrem kleinen Wert,:

Wir hatten viel zu tun. Und wir haben viel getan. Jetzt können wir darüber reden. Bitte informieren Sie sich. Sie werden sehen, es waren drei erfolgreiche Jahre.

Text A, mit einem extrem großen Wert,:

Nach der Liquidierung des studentischen Mentorenprogramms und der Streichung der Mittelzuwendung für das Zentrum für politische Bildung holt jetzt der Senat in Reaktion auf CDU-Expertise und Brief der 28 in konzertierter Aktion mit den Rechtskräften vom Bund Freiheit der Wissenschaft zum Schlag gegen die demokratischen Mitbestimmungspositionen der Studenten in den Fachbereichs- und Institutsräten aus.

Von Interesse ist es festzustellen, daß alle Texte nach dem Merkmal Wortlänge zu der Gruppe der Schriftsteller bei Fucks gehören, nach dem Merkmal Satzlänge sowohl zu den Dichtern als auch zu den Schriftstellern. Für Texte aus der letzten Gruppe hat Fucks festgestellt, für manche Fälle träfe die Erklärung zu, daß "die Schriftsteller in ihren Fachsprachen schreiben, deren Wörter meist überdurchschnittlich lang sind, und daß Hand in Hand damit eine Tendenz zu längeren Sätzen gehe."⁷ In den Flugblättern läßt sich diese Tendenz allerdings ebenso feststellen, vgl. die häufig zu findenden Wörter wie Mitbestimmungsrecht, Hochschulrahmengesetz, Prüfungs- und Berufungsangelegenheiten, Mitbestimmungspositionen, drittelparitätisch.

Bei der Wahlwerbung (Tabelle 2) fällt auf, daß die Wortlänge in den meisten Texten rund 1,9 beträgt. Vergleicht man in dieser Gruppe die drei SPD-Politiker⁸ Schmidt, Arendt und Brandt, so ergeben sich folgende Daten:

Autor	Textlänge		\bar{i}	\bar{j}	$\bar{i} \cdot \bar{j}$
	Sätze	Wörter			
Schmidt	24	588	1,949	24,500	47,748
Arendt	39	604	2,247	15,487	34,793
Brandt	32	375	1,947	11,719	22,812

Tabelle 5 : Mittlere Wortlänge \bar{i} , mittlere Satzlänge \bar{j} ,
 ξ -Werte $\bar{i} \cdot \bar{j}$ für SPD-Politiker.

Der Wortlänge nach sind Schmidt und Brandt beinahe identisch, Arendt verwendet dagegen aber Wörter, die im Durchschnitt 15 % länger sind als bei Brandt und Schmidt. Diese gebrauchen Wörter, deren Durchschnittslänge sehr nahe an die von Einstein⁹, *Evolution der Physik*: $\bar{i} = 1,929$, Ostwald, *Farbkunde*: $\bar{i} = 1,969$ und Du Bois Reymond, *Über eine Akademie der deutschen Sprache*: $\bar{i} = 1,939$ kommt. Auch der Satzlänge nach kommt Brandt dem Mittelwert $\bar{j} = 11,181$ von Du Bois Reymond sehr nahe, ein Wert, der in der Gruppe der Schriftsteller extrem niedrig ist und auch bedeutend unter dem Mittelwert 16 bei Eggers liegt. Arendt liegt diesem Mittelwert sehr nahe, während Schmidt mit der mittleren Satzlänge $\bar{j} = 24,500$ dem Mittelwert der deutschen Erzählprosa, ermittelt durch Arens, $\bar{j} = 25,40$ (vgl. Tabelle 3) nahe kommt.

Folgender Vergleich mit Brandts Rede vor der Vollversammlung der Vereinten Nationen zeigt jedoch, daß sich seine mittlere Satzlänge mit dem Durchschnitt der von Fucks untersuchten 50 Texte fast genau deckt. Vom Anfang der Rede wurde ein Abschnitt von 388 Wörtern ausgezählt mit dem Ergebnis $\bar{i} = 1,935$ und $\bar{j} = 21,556$. Der Mittelwert \bar{j} für sämtliche Texte von Fucks ist $\bar{j} = 21,55$, s. Tabelle 3. Aus Brandts Rede sehen wir, daß die Wortlänge mit 1,935 Silben im Durchschnitt seiner Wortlänge in der Wahlwerbung sehr nahe kommt. Weitere Stichproben müssen analysiert werden, um die Frage zu beantworten, ob wir es hier mit einer autoren-spezifischen Wortlänge zu tun haben.

Zusammenfassend kann festgestellt werden, daß die beiden Textgruppen sich vorwiegend durch den Parameter Satzlänge (56 %) unterscheiden, ebenso wie durch die Werte \bar{i}, \bar{j} . Die Wortlänge unterscheidet sich nur um 8 %. In beiden Gruppen liegt die Wortlänge aber höher als der Mittelwert für die deutsche Prosa.

Wir analysieren nun die beiden Textgruppen anhand der Parameter Adjektive je Wort (A/W), Adjektive je Verb (A/V) und Adjektive je Satz (A/S).

Flugblätter	A/W	A/V . 100	A/S
A	0,056	65,573	1,666
B	0,036	25,960	0,774
C	0,048	35,922	1,370
D	0,052	26,506	1,000
E	0,084	70,373	1,900
F	0,078	10,074	1,933
	0,354	234,408	8,643
	M = 0,059	39,067	1,441

Tabelle 6 : Mittlere Quotientenwerte für die Flugblätter.

Adjektiv-Wort-Quotient, Adjektiv-Verb-Quotient,
Adjektiv-Satz-Quotient.

Die Flugblätter A (Asta-Info. Behörde plant Abschaffung der Drittelparität) und E (Sofortige Aufnahme des Streiks) zeigen in Vergleich mit anderen einen bedeutend höheren Adjektiv-Verb-Quotienten. Es fällt auf, daß auch die anderen Quotienten bei E erheblich höher sind als bei den meisten anderen.

Für die Adjektiv-Verb-Quotienten als Differenzierungsmaßstab für Textarten können die Mittelwerte von David D. Boder¹⁰ herangezogen werden: 11,2 für Dramentexte, 20,0 für Romane und 75,5 für wissenschaftliche Texte. Man darf natürlich nicht vergessen, daß andere Samples diese Zahlen vielleicht ändern könnten, wir können aber auch schon mit deren Hilfe gewissen Tendenzen feststellen. Vor allem muß im Auge behalten werden, daß Boders Angaben für Mittelwerte gelten.¹¹ Die Flugblätter stehen mit $A/V.100 = 39,067$ zwischen den Romanen und wissenschaftlichen Texten.

Dieselben Parameter ergeben für die Wahlwerbung folgende Werte:

Wahlwerbung	A/W	A/V.100	A/S
1.	0,048	30,337	0,600
2.	0,079	68,292	1,000
3.	0,026	30,000	0,625
4.	0,044	50,000	0,813
5.	0,056	75,000	0,872
6.	0,048	37,500	0,563
	0,301	291,129	4,473
	M = 0,050	48,522	0,746

Tabelle 7 : Mittlere Quotientenwerte für die Wahlwerbung.
Adjektiv-Wort-Quotient, Adjektiv-Verb-Quotient,
Adjektiv-Satz-Quotient.

Die Wahlwerbung zeigt in Text 2 und 5 bei A/V hohe Werte. Der hohe Adjektiv-Verb-Quotient in dieser Art der Texte (Text 2: Wir führen das Gespräch mit dem Wähler; Text 5: Das muß jedermann wissen über die Rentenversicherung, W. Arendt) scheint die Klassifizierungsrichtung von Boder zu bestätigen.

Bei der Wahlwerbung fällt der niedrige Adjektiv-Satz-Quotient auf: der höchste Wert ist 1,000 in Text 2, der niedrigste 0,563 in Text 6.

Einen Vergleich der drei Quotienten bei den SPD-Politikern ermöglicht Tabelle 8 (vgl. hierzu auch Tabelle 5).

Autor	A/W	A/V.100	A/S
Schmidt	0,026	30,000	0,625
Arendt	0,056	75,000	0,872
Brandt	0,048	37,500	0,563

Tabelle 8 : Mittlere Quotientenwerte für SPD-Politiker.

Bei diesen Parametern ist der Unterschied zwischen den drei Politikern größer als bei der Wort- und Satzlänge. Arendt verwendet etwas mehr als zweimal so viel Adjektive je Wort, zweieinhalbmals so viele Adjektive je Verb und rund die Hälfte mehr Adjektive je Satz als Schmidt. Brandt verwendet rund 90 % mehr Adjektive je Wort, 25 % mehr Adjektive je Verb, jedoch rund 40 % weniger Adjektive je Satz als Schmidt. Durch diese Parameter läßt sich der intuitiv erfaßte Unterschied zwischen den Texten, wie wir gesehen haben, in weitere objektivierbare Relationen bringen.

Der Vergleich der beiden Textgruppen ergibt folgendes. Der Adjektiv-Wort-Quotient ist in den beiden Textgruppen sehr ähnlich, sie unterscheiden sich jedoch in den Werten des Adjektiv-Verb-Quotients und des Adjektiv-Satz-Quotients. Die Texte der Wahlwerbung verwenden etwa 25 % mehr Adjektive je Verb als die Flugblätter. Die Flugblätter haben aber rund 93 % mehr Adjektive je Satz als die Wahlwerbung. Der markanteste Unterschied zwischen den Textgruppen zeigt sich in diesem Parameter, nicht nur in den Mittelwerten, sondern auch, wenn wir die Streuung berücksichtigen. Die Flugblätter verwenden beinahe zweimal so viel Adjektive je Satz wie die Wahlwerbung. Das Auffallende dabei ist jedoch, daß die Wahlwerbung so wenig Adjektive verwendet: $A/S = 0,746$. Weitere Untersuchungen sind erforderlich, um die Reichweite dieser Tendenz zu prüfen.

Zur quantitativen Stilanalyse müssen zahlreiche Parameter herangezogen werden. Schon die hier angegebenen haben allerdings gezeigt, in welcher Weise die quantitativ feststellbaren Struktureigenschaften Texte charakterisieren können. Die Stilindikatoren, wodurch sich die beiden Textgruppen der Werbetexte unterscheiden, die wir als Flugblätter und Wahlwerbung gekennzeichnet hatten, sind vor allem die mittlere Satzlänge, die mittlere Wortlänge und mittlere Satzlänge zugleich, der Adjektiv-Verb-Quotient und der Adjektiv-Satz-Quotient.

Weitere Untersuchungen auf diesem Gebiete sind für die Textlinguistik wichtig, denn "Die Stilcharakteristiken, die durch quantitativ faßbare Struktureigenschaften der Texte definiert sind, können sprach-, gattungs-, autor- oder werkspezifisch sein."¹² Ferner: Sprache ist nicht nur ein Ausdrucks- und Kommunikationsmittel des Menschen, sondern auch immer ein Identifizierungsinstrument.¹³ Die Methoden der Stilanalyse auf mathematisch-statistischer Grundlage können in sehr großem Maße be-

hilflich sein, daß Funktionieren dieses Instruments zu beleuchten.

Anmerkungen

- 1 Erörterungen der Methodik und der Anwendungsgebiete bei Els Oksaar, Stilstatistik und Textanalyse. In: Festschrift für Hans Eggers zum 65. Geburtstag, Tübingen 1972, S. 630 - 648, und Els Oksaar, Sprache als Problem und Werkzeug des Juristen. In: Archiv für Rechts- und Sozialphilosophie 53, 1967, S. 91 - 130. Es ist sehr wichtig, im Auge zu behalten, daß die Textanalyse mit quantitativen Methoden von einem im Vergleich mit den qualitativen Methoden anderen Stilbegriff ausgeht. Vgl. Wilhelm Fucks, Mathematische Analyse des literarischen Stils. In: Studium Generale 6, 1953, S. 507, der Stil als "Gesamtheit aller quantitativ faßbaren Gegebenheiten in der formalen Struktur eines Textes" versteht.
- 2 Für technische Hilfe bei den Auszählungen und Rechenarbeiten danke ich Fräulein Dagmar Baumhögger und Herrn Hans Ludwig Wirkner.
- 3 Wilhelm Fucks, Unterschied des Prosastils von Dichtern und anderen Schriftstellern, Ein Beispiel mathematischer Stilanalyse. In: Sprachforum 1, 1955, S. 236 ff.
- 4 Hans Arens, Verborgene Ordnung. Beihefte zur Zeitschrift "Wirkendes Wort", Düsseldorf 1965, S. 44.
Hans Eggers, Wandel im deutschen Satzbau. In: Der Deutschunterricht 13, Heft 5, 1961, S. 47 ff.
- 5 Wilhelm Fucks [Anm. 3], S. 241.
- 6 Wilhelm Fucks, ebd., S. 243.
- 7 Wilhelm Fucks, ebd., S. 243.
- 8 Analysen verschiedener Texte von Politikern aller Parteien werde ich in einem anderen Zusammenhang veröffentlichen.
- 9 Daten nach Wilhelm Fucks [Anm. 3], S. 236 u. 243.
- 10 "The Adjective-Verb Quotient: A contribution to the Psychology of Language". In: Psychological Record 3, 1940, S. 310 - 343.
- 11 Wenn Streuungen berücksichtigt werden, ergibt sich, je nach dem persönlichen Stil des Autors, immer ein anderes Bild. Es nimmt daher nicht Wunder, wenn Hardi Fischer, Entwicklung und Beurteilung des Stils. In: Mathematik und Dichtung, München 1965, S. 180, bei Schriftstellern wie Maria Waser einen extrem hohen Adjektiv-Verb Quotienten feststellt ($AN : 100 = 167$) und wir es ja nicht mit einem wissenschaftlichen Text zu tun haben. — Die

Daten von Fischer, die Texte von Dichtern wie Schwab, von Ebner-Eschenbach, Herder, Pestalozzi, Raabe u.a. als Grundlage haben, ergeben nach unseren Rechnungen den Mittelwert $A/W \cdot 100 = 52,2$.

- 12 Wilhelm Fucks und Josef Lauter, Mathematische Analyse des literarischen Stils. In: Mathematik und Dichtung. München 1965, S. 109.
- 13 Vgl. hierzu Els Oksaar, Sprache als Problem und Werkzeug des Juristen [Anm. 1], S. 91 ff.